

Causal Inference in HIV Prevention Trials

Michael Hudgens, PhD
University of North Carolina
Department of Biostatistics
Center for AIDS Research

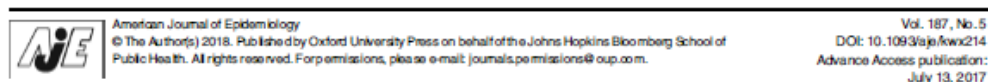
Causal Inference in HIV Prevention Trials

- Suppose an exposure (or treatment) associated with rate of HIV acquisition in a study
- Eg, contraceptive associated ($p < .05!$) with lower HIV incid.
- When can we conclude that such an association is due to exposure actually having an **effect** on HIV infection?
- Causal inference methods provide a formal statistical framework for deducing causal claims from data

Causal Inference

- Long history in philosophy
- Early statistical work by Neyman in 1920s, then important work by Rubin 1970s and Robins 1980s lead to modern era where commonly used approach in numerous disciplines
- Important (arguably fundamental) in epidemiology, econometrics, comparative effectiveness research, implementation science, public policy, and many other areas of research
- Books, journals, conferences, courses, software, blogs

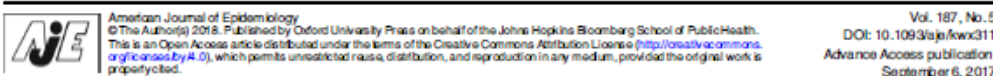
Causal Inference in Epidemiology



Practice of Epidemiology

Propensity Score–Based Methods in Comparative Effectiveness Research on Coronary Artery Disease

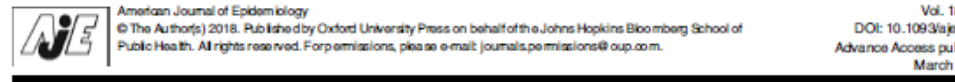
Alexandra G. Ellis, Thomas A. Trikalinos, Benjamin S. Wessler, John B. Wong, and Issa J. Dahabreh*



Practice of Epidemiology

Analysis of Longitudinal Studies With Repeated Outcome Measures: Adjusting for Time-Dependent Confounding Using Conventional Methods

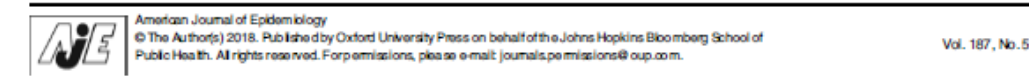
Ruth H. Keogh*, Rhian M. Daniel, Tyler J. VanderWeele, and Stijn Vansteelandt



Practice of Epidemiology

A Note on G-Estimation of Causal Risk Ratios

Oliver Dukes* and Stijn Vansteelandt



Research Letter

INVERSE PROBABILITY WEIGHTS FOR THE ANALYSIS OF POLYTOMOUS OUTCOMES

HPTN is in the causal inference business

- *The HIV Prevention Trials Network is a worldwide collaborative clinical trials network that develops and tests the safety and efficacy of interventions designed to prevent the transmission of HIV*
- HPTN : learn about the **causal effects** of interventions on safety and HIV transmission outcomes

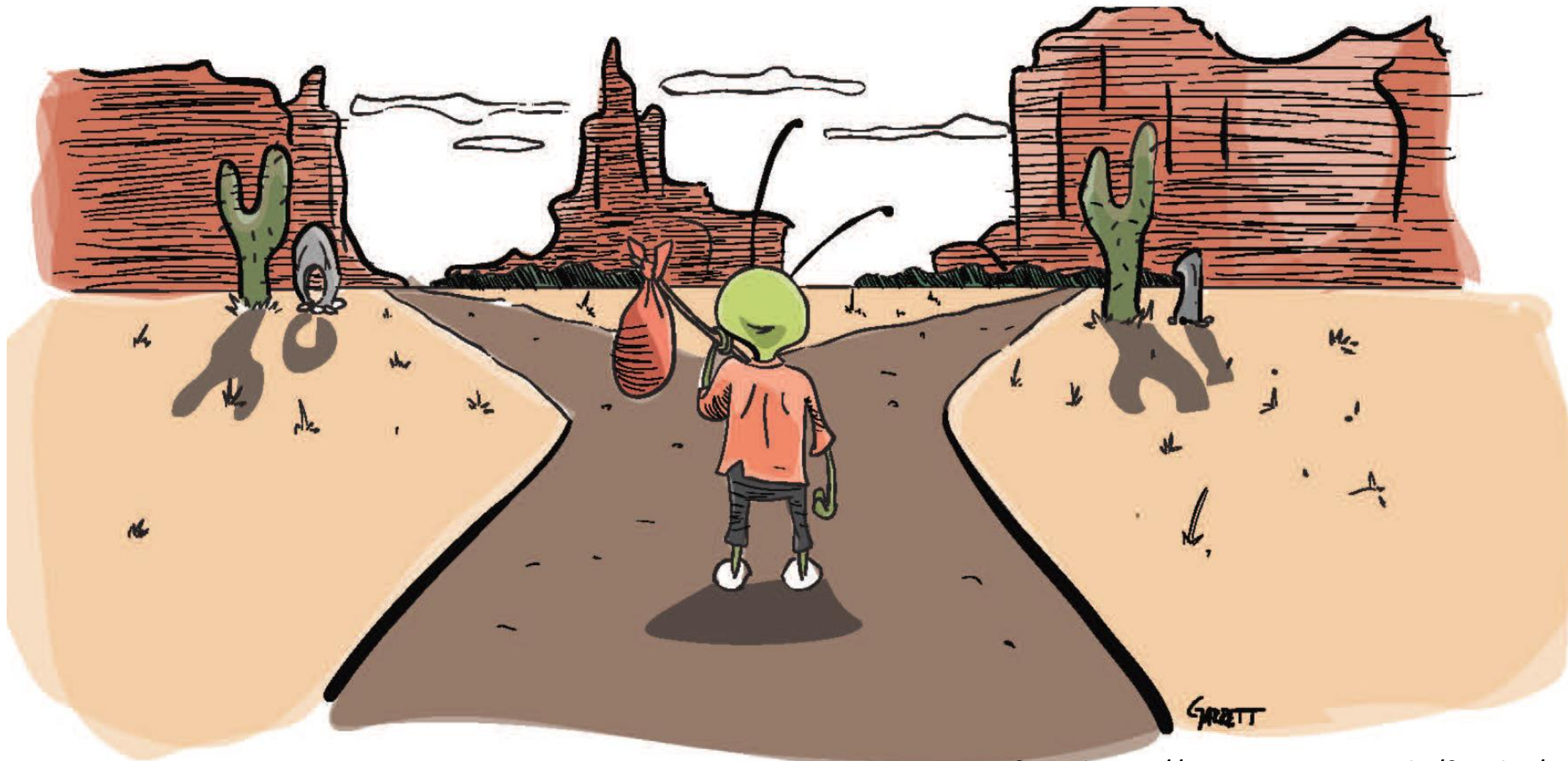
Outline

- A gentle introduction to drawing causal inference using data from biomedical studies
- Cover basic ideas like potential outcomes, counterfactuals, confounding, propensity score, etc
- Examples from HIV prevention studies



Potential Outcomes

- Binary treatment (exposure) with values $A = 0,1$
- For example, $A = 1$ treatment, $A = 0$ no treatment (control)
- Consider two **potential outcomes**
 - $Y(1)$ if individual receives treatment
 - $Y(0)$ if individual receives control
- Eg, $Y(a) = 1$ if individual HIV+ after 5 years for $a = 0,1$



Causal Effects

- If $Y(1) = Y(0)$, then treatment has no (causal) effect
- Otherwise, treatment has an effect
- Typically we only observe $Y(1)$ or $Y(0)$ but not both
- Eg if individual receives treatment $A = 1$, we observe $Y(1)$, and $Y(0)$ becomes **counterfactual** (missing)
- Thus estimating effect at individual level generally not possible

Causal Effects

- Consider estimating effects at the population level such as **average treatment effect**

$$E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$$

- Interpretation:
 - Average individual-level effect of treatment
 - Difference in average outcome if everyone receives treatment versus if no one receives treatment (two counterfactual scenarios)

Causal Effects

- Consider estimating effects at the population level such as **average treatment effect**

$$E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$$

- For binary outcome, consider **causal RD or RR**

$$\Pr[Y(1) = 1] - \Pr[Y(0) = 1] \text{ or } \Pr[Y(1) = 1] / \Pr[Y(0) = 1]$$

Randomized experiments

- Suppose we randomly assign individuals to treatment
- Treatment assignment A is **independent** of $Y(1)$ and $Y(0)$
 $A \perp Y(1), Y(0)$
- Denote the observed outcome by Y , e.g., $Y = Y(1)$ if $A = 1$
- Then $E[Y(1)] = E[Y(1)|A = 1] = E[Y|A = 1]$
- Thus can estimate $E[Y(1)]$ by the mean outcome in individuals randomized to treatment; likewise for $E[Y(0)]$

Randomized experiments

- Therefore we can estimate the average treatment effect

$$E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$$

by difference in sample means (t-test), or by fitting a simple linear regression model

$$E[Y|A] = b_0 + b_1 A$$

- Upshot: **causal inference straightforward in randomized trials**

Randomized experiments

- Proviso: Randomized experiments often include issues like drop-out, non-compliance, measurement error, etc

The Journal of Infectious Diseases

MAJOR ARTICLE



Pharmacological Measures of Treatment Adherence and Risk of HIV Infection in the VOICE Study

James Y. Dai,^{1,2} Craig W. Hendrix,³ Barbra A. Richardson,² Cliff Kelly,¹ Mark Marzinke,³ Z. Mike Chirenje,⁴ Jeanne M. Marrazzo,² and Elizabeth R. Brown^{1,2}

¹Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, and ²University of Washington, Seattle; ³School of Medicine, Johns Hopkins University, Baltimore, Maryland; and

⁴University of Zimbabwe–University of California, San Francisco, Research Programme, Harare

(See the editorial commentary by Vermund and Walker on pages 329–31.)

Randomized experiments

- In addition, we may be interested in secondary analyses of trial data where the treatment/exposure was not randomized, or in mediation (causal pathways)

Kilburn K et al. *Journal of the International AIDS Society* 2018, **21**(S1):e25043
<http://onlinelibrary.wiley.com/doi/10.1002/jia2.25043/full> | <https://doi.org/10.1002/jia2.25043>

RESEARCH ARTICLE

Conditional cash transfers and the reduction in partner violence for young women: an investigation of causal pathways using evidence from a randomized experiment in South Africa (HPTN 068)

Kelly N Kilburn^{1,§}, Audrey Pettifor^{1,2}, Jessie K Edwards¹, Amanda Selin¹, Rhian Twine², Catherine MacPhail^{3,4}, Ryan Wagner^{2,5}, James P Hughes^{6,7}, Jing Wang⁶ and Kathleen Kahn^{2,8}

Observational studies

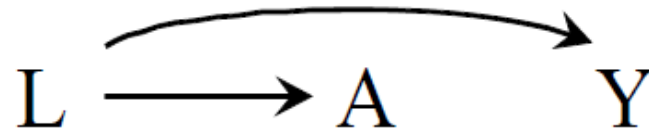
- Consider study where treatment received is not randomized, so we are no longer willing to assume

$$A \perp Y(1), Y(0)$$

- Eg, A = contraceptive, Y = incident HIV infection
- Women who are more sexually active may be more likely to use contraceptive ($A = 1$) and may also be more likely, regardless of treatment, to acquire HIV ($Y(1) = Y(0) = 1$)

Observational studies

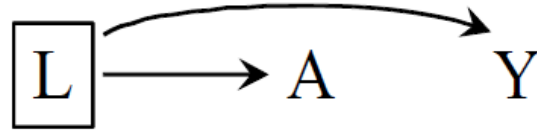
- Directed acyclic graphs (DAGs) often used in causal inference to depict assumptions
- L sexual behavior, A oral contraceptive, Y HIV infection



- L confounds the association between A and Y
- In general, A will not be independent of $Y(0)$, $Y(1)$

Observational studies

- However, suppose we only consider women with the same sexual behavior

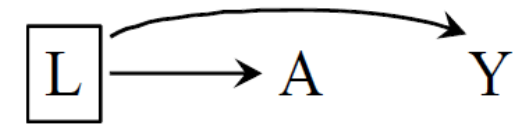


- We might be willing to assume that $A \perp Y(1), Y(0) \mid L$
- In other words, within strata defined by L , as-if randomized trial

Observational studies

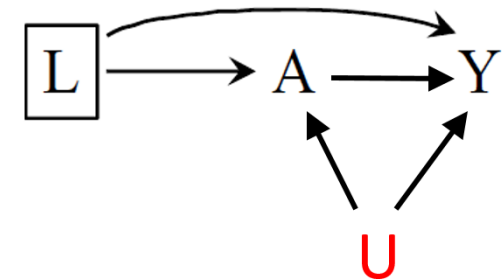
- Key assumption

$$A \perp Y(1), Y(0) \mid L$$



might be more plausible if condition on additional covariates

- Conditional on age, race, ethnicity, sexual behavior, education, etc, women who select OC similar to women who do not
- **Conditional exchangeability**, or **no unmeasured confounders** assumption



Observational studies

- Conditional exchangeability key assumption underlying most causal inference methods
- Within strata defined by covariates $L = (L1, L2, \dots)$, as-if randomized trial
- Recall: causal inference in randomized studies easy
- Implication: estimate causal effects within strata, then average estimates across strata
- Essential idea between many **matching** methods and **standardization/g-formula**

Can't I just do multivariate regression?

- What if we fit this model?

$$E[Y|A, L] = b_0 + b_1 A + b_2 L$$

- Is $\hat{b}_1$ valid estimate of the causal effect?
- Yes, if model correct
- But model supposes effect of treatment same for all individuals, which will be implausible in many settings

Propensity score

- **Propensity score** is conditional probability individual receives treatment given covariates $e(L) = \Pr[A = 1|L]$
- Important result. Conditional exchangeability implies
$$A \perp Y(1), Y(0) \mid e(L)$$
- That is, if it is sufficient to adjust/control for L , then it is sufficient to match/stratify by $e(L)$
- Advantageous because $e(L)$ is just a scalar

Propensity score

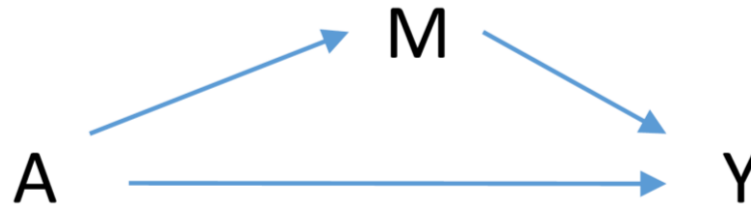
- Propensity score generally unknown in observational studies
- Estimate via logistic regression $\text{logit}(\Pr[A = 1]) = a_1 + a_2 L$
- Stratify or match individuals based on estimate propensity scores
- Estimate causal effects within strata/matches, then average over effect estimates

Inverse Probability Weighting

- Another common approach using propensity scores
- Estimate via logistic regression $\text{logit}(\Pr[A = 1]) = a_1 + a_2 L$
- Estimate $E[Y(1)]$ by weighted average of Y among treated individuals $A = 1$ with weights $1/e(L)$
- Similarly for $E[Y(0)]$
- Creates pseudo-population where no confounding

Other areas of causal inference

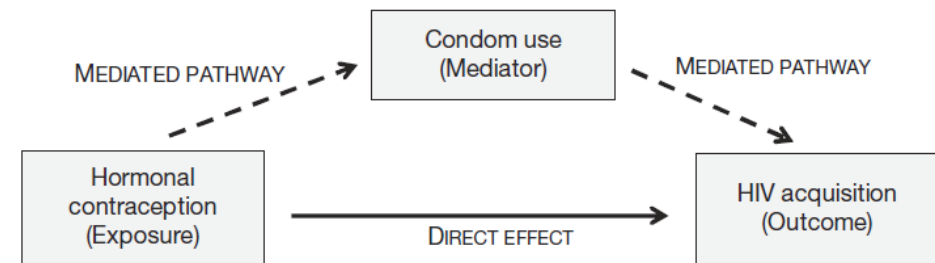
- Mediation



Assessing the effect of hormonal contraception on HIV acquisition in observational data: challenges and recommended analytic approaches

Chelsea B. Polis^{a,*}, Daniel Westreich^{b,c,*}, Jennifer E. Balkus^{d,e,*},
Renee Heffron^{e,*}, participants of the 2013 HC-HIV
Observational Analysis Meeting

AIDS 2013, **27** (Suppl 1):S35–S43



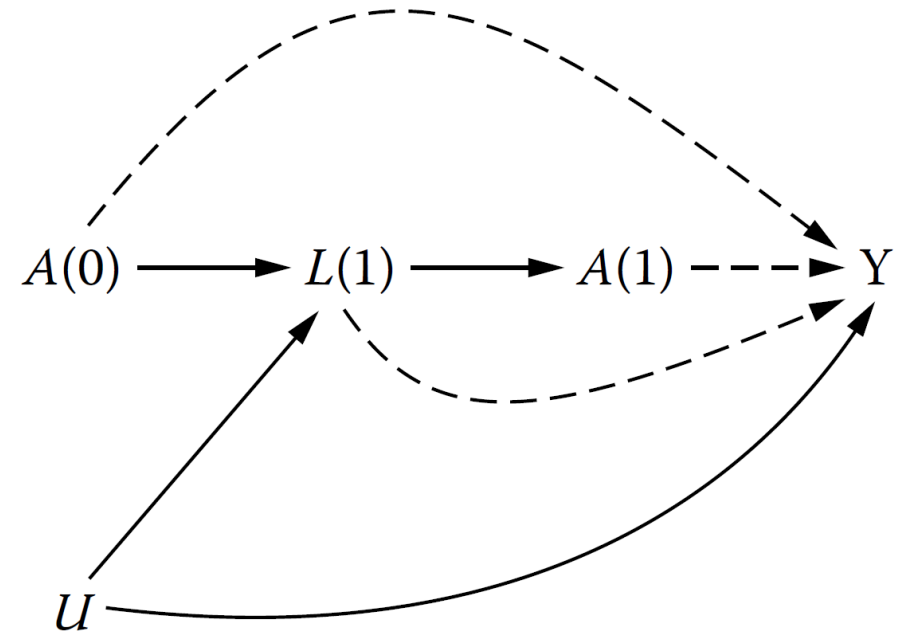
Other areas of causal inference

- Mediation
- Time-varying exposures

$A(0)$ ART at visit 0

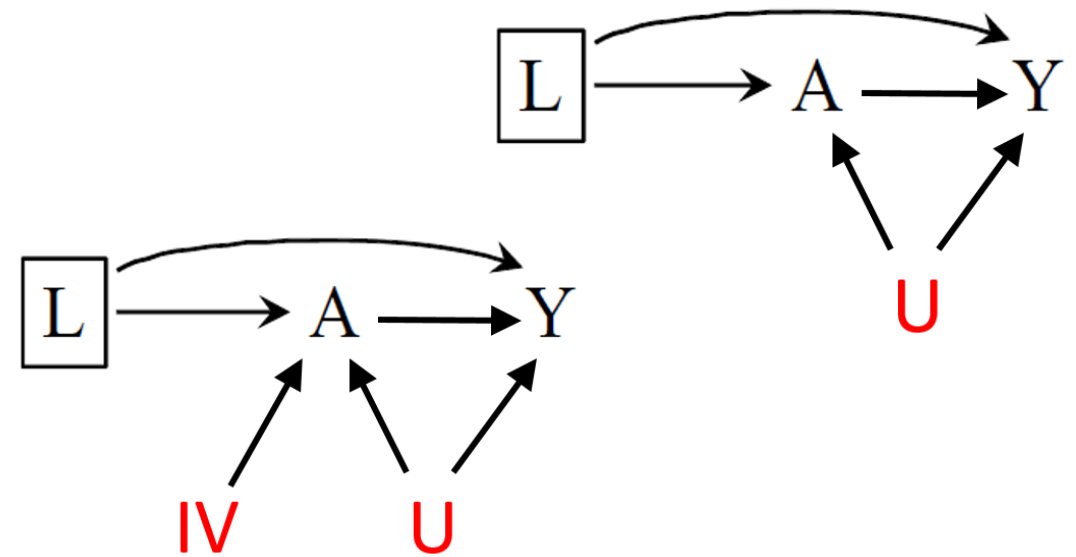
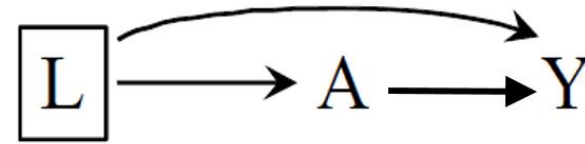
$L(1)$ CD4 at visit 1

$A(1)$ ART at visit 1



Other areas of causal inference

- Mediation
- Time-varying exposures
- Instrumental variables:
 - $IV - Y$ unconfounded, and
 - IV has an effect on Y only via A



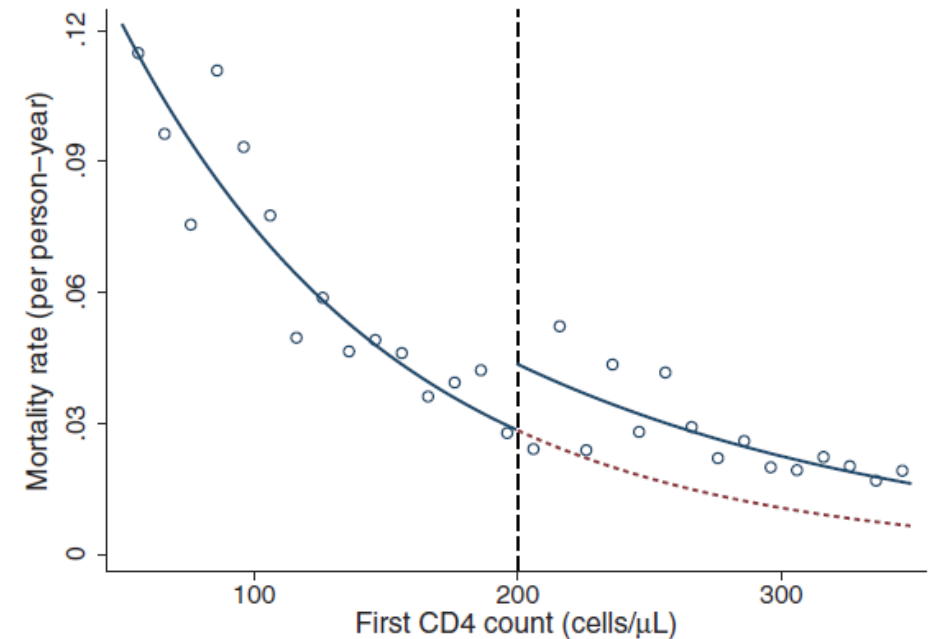
Other areas of causal inference

- Mediation
- Time-varying exposures
- Instrumental variables
- Regression discontinuity designs

Regression Discontinuity Designs in Epidemiology *Causal Inference Without Randomized Trials*

Jacob Bor,^{a,b,c} Ellen Moscoe,^c Portia Mutevedzi,^b Marie-Louise Newell,^{b,d} and Till Bärnighausen^{b,c}

Epidemiology • Volume 25, Number 5, September 2014



Other areas of causal inference

- Mediation
- Time-varying exposures
- Instrumental variables
- Regression discontinuity designs
- Principal stratification

$D(1)$ indicator if person would adhere when assigned treatment

The Journal of Infectious Diseases

MAJOR ARTICLE

 IDSA
Infectious Diseases Society of America

 hivma
hiv medicine association

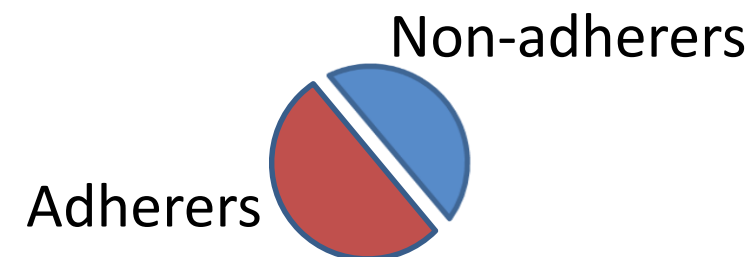
 OXFORD

Pharmacological Measures of Treatment Adherence and Risk of HIV Infection in the VOICE Study

James Y. Dai,^{1,2} Craig W. Hendrix,³ Barbra A. Richardson,² Cliff Kelly,¹ Mark Marzinko,² Z. Mike Chirenje,⁴ Jeanne M. Marrazzo,² and Elizabeth R. Brown^{1,2}

¹Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, and ²University of Washington, Seattle; ³School of Medicine, Johns Hopkins University, Baltimore, Maryland; and ⁴University of Zimbabwe–University of California, San Francisco, Research Programme, Harare

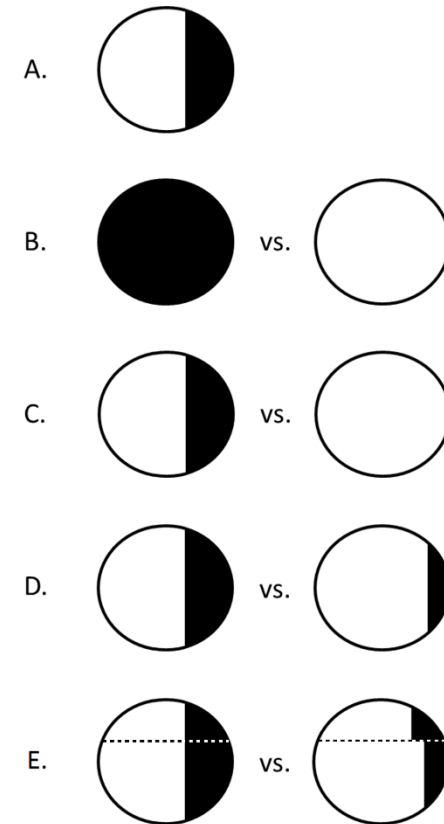
(See the editorial commentary by Vermund and Walker on pages 329–31.)



Casual Effects (Revisited)

- Recall causal effect definition: all individuals exposed vs no individuals exposed
- Other quantities may be more relevant from policy/public health perspective
- Effect of treatment in treated
- E.g., smoking

From Patients to Policy
Population Intervention Effects in Epidemiology
Daniel Westreich

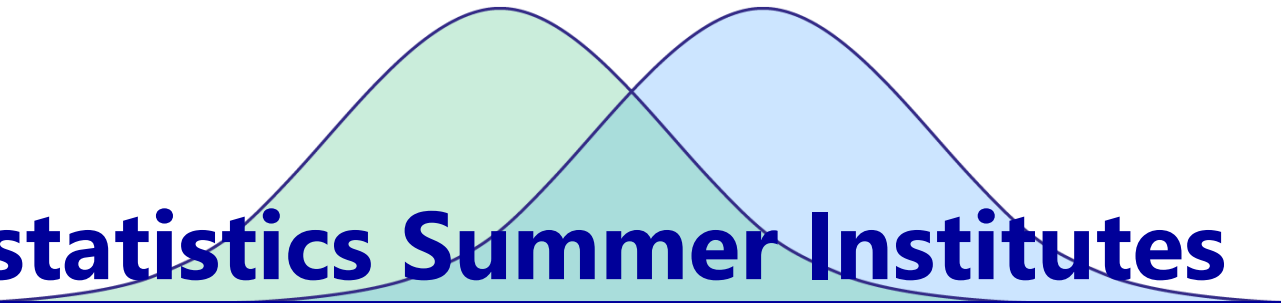


Conclusion

- Causal inference central to mission of HPTN
- Straightforward in randomized studies w/ perfect compliance, etc
- Specialized statistical methods for observational studies
 - Matching, stratification/standardization
 - Inverse probability weighting
- Can utilize these methods in HPTN trials with imperfect compliance, LTFU, exposures that were not randomized, mediation analysis, etc.

Causal Inference Resources

- Books ([Hernan and Robins](#), [Imbens and Rubin](#), ...)
- Journals ([Journal of Causal Inference](#))
- Software ([SAS Proc Causaltrt](#), R packages, ...)
- Conferences ([Atlantic CI Conference](#), [EuroCIM](#), ...)
- Short courses ([UW SIS MID](#) July 23-25, [Harvard](#) June 4-8)



UW Biostatistics Summer Institutes

Seattle

UNIVERSITY OF WASHINGTON | SCHOOL OF PUBLIC HEALTH | DEPARTMENT OF BIOSTATISTICS

July 9-27, 2018

4
Institutes

- Statistical Genetics
- Statistics and Modeling in Infectious Diseases
- Statistics for Clinical Research
- Statistics in Big Data

57
Short
Courses

84
Instructors
who are leading
educators and
researchers

900+
Participants
in 2017

For more information or to apply for a scholarship: biostat.washington.edu/suminst

Causal Inference in HIV Prevention Trials

Michael Hudgens, PhD
University of North Carolina
Department of Biostatistics
Center for AIDS Research